

## ÜBUNGEN ZUR WAHRSCHEINLICHKEITSRECHNUNG UND STATISTIK FÜR BIOLOGEN

## Blatt 13

**Aufgabe 1** Im Jahr 2003 wurden im Sichote-Alin-Naturreservat einige männliche, ausgewachsene Sibirische Tiger gewogen und ihr Nachwuchs markiert. 2011 wurden vier der damals markierten männlichen, inzwischen ausgewachsenen Nachkommen aufgespürt und ebenfalls gewogen. In der folgenden Tabelle stehen jeweils die Gewichte von Vater und Sohn untereinander.

Gewicht Vater (2003) in kg	190	220	220	250
Gewicht Sohn (2011) in kg	190	230	240	260

- (a) Anhand der Daten soll untersucht werden, ob sich das mittlere Gewicht der männlichen Tiger im Reservat von der vorherigen zur jetzigen Generation verändert hat. Führen Sie einen geeigneten Test mit Signifikanzniveau  $\alpha = 0.05$  durch. Stellen Sie den Rechenweg nachvollziehbar dar und formulieren Sie Ihr Ergebnis in einem adäquaten Satz.
- (b) Um wieviele kg ist ein 2011 im Reservat lebendes ausgewachsenes Tigermännchen im Durchschnitt leichter/schwerer als sein Vater? Berechnen Sie für diesen Mittelwert ein 95%-Konfidenzintervall.

**Aufgabe 1 – online-Version** Im Jahr 2003 wurden im Sichote-Alin-Naturreservat einige männliche, ausgewachsene Sibirische Tiger gewogen und ihr Nachwuchs markiert. 2011 wurden vier der damals markierten männlichen, inzwischen ausgewachsenen Nachkommen aufgespürt und ebenfalls gewogen. In der folgenden Tabelle stehen jeweils die Gewichte von Vater und Sohn untereinander.

Gewicht Vater (2003) in kg	190	220	220	250
Gewicht Sohn (2011) in kg	190	230	240	260

Anhand der Daten soll untersucht werden, ob sich das mittlere Gewicht der männlichen Tiger im Reservat von der vorherigen zur jetzigen Generation verändert hat. Sie wolle einen geeigneten Test mit Signifikanzniveau  $\alpha = 0.05$  durchführen.

- (a) Welcher Test kommt hier am ehesten in Frage?
- der ungepaarte einseitige zwei-Stichproben-t-Test mit gleichen Varianzen
  - der ungepaarte zweiseitige Welch-t-Test
  - der ungepaarte einseitige Welch-t-Test
  - der gepaarte zweiseitige zwei-Stichproben-t-Test
  - der gepaarte einseitige zwei-Stichproben-t-Test
- (b) Wie groß muss der Betrag des t-Werts mindestens sein, damit sie die Nullhypothese verwerfen können? (Angabe mit zwei Nachkommastellen.)
- (c) Angenommen, der p-Wert beträgt 0.09. Wie sollten Sie dann das Ergebnis des Tests zusammenfassen?
- Die Nullhypothese, dass es keinen signifikanten Unterschied im mittleren Gewicht von Vätern und Söhnen gibt, kann nicht verworfen werden.
  - Der Unterschied zwischen dem mittleren Gewicht der Väter und dem der Söhne ist nicht signifikant.

- Wir verwerfen die Nullhypothese, dass sich das Gewicht von Vater zum Sohn im Durchschnitt ändert.
  - Das durchschnittliche Gewicht bleibt von einer Generation zur nächsten gleich.
  - Die Söhne haben ein signifikant anderes Gewicht als die Väter.
- (d) Um wieviele kg ist ein 2011 im Reservat lebendes ausgewachsenes Tigermännchen im Durchschnitt leichter oder schwerer als sein Vater?
- (e) Berechnen Sie den Standardfehler dieses Mittelwerts. (Angabe mit drei Nachkommastellen.)
- (f) Berechnen Sie für diesen Mittelwert ein 95%-Konfidenzintervall. (Angabe der Intervallgrenzen mit drei Nachkommastellen.)

## Aufgabe 2

- (a) Angenommen, die Mittelwerte in den in der Tabelle in Aufgabe 1 gegebenen Stichproben stimmen genau mit den entsprechenden Mittelwerten in den Populationen der Jahre 2003 und 2011 überein. Wieviele markierte Tigersöhne von verschiedenen Vätern müssten Sie dann 2011 aufspüren, damit der Unterschied mit fünfundneunzigprozentiger Wahrscheinlichkeit auf einem Signifikanzniveau von  $\alpha = 0.01$  statistisch nachgewiesen werden kann?
- (b) Die Stichprobe wurde folgendermaßen erhoben: Unter allen männlichen geschlechtsreifen Tigern im Jahr 2003 wurden einige Tiger mit Hilfe einer Liste, in der alle Großkatzen des Nationalparks erfasst sind, rein zufällig ausgewählt. Vom Nachwuchs dieser Tiger wurde dann jeweils ein männliches Tier ausgewählt. (Tiger, die keinen männlichen Nachwuchs hatten, konnten natürlich nicht berücksichtigt werden.) Diskutieren Sie für die Väter sowie für die Söhne, ob es sich jeweils um eine repräsentative Stichprobe für die männlichen Tiger ihrer Generation im Nationalpark handelt.

**Aufgabe 3** Die Therapie der Influenza mit Neuraminidasehemmern wie Zanamivir oder Oseltamivir sollte innerhalb von 48 Stunden nach Auftreten der ersten Krankheitssymptome beginnen. Da eine laborbasierte Diagnostik zu lange dauern würde, ist es notwendig, kurzfristig eine gesicherte Diagnose zu stellen. Hier bietet sich der Einsatz eines Influenza-Schnelltests an. Einer dieser Schnelltests hat eine Sensitivität von 80% (d.h. er schlägt bei 80% der Infizierten an) und eine Spezifität von 85% (d.h. er schlägt bei 15% der Nicht-Infizierten an). Diese Werte hängen von den genauen Umständen ab. Beispielsweise kann Nahrungsaufnahme kurz vor dem Schnelltest bei einem Rachenabstrich zu negativen Ergebnissen führen, da dann die Anzahl der Viren im Rachenraum verringert ist. Frau Meyer liest einen Bericht über die Krankheit in der Zeitung, macht sich Sorgen und lässt am nächsten Tag einen Influenza-Schnelltest durchführen. Das Testergebnis ist negativ. Wie wahrscheinlich ist es, dass Frau Meyer dennoch mit Influenza infiziert ist? Sehen Sie Frau Meyer hierfür als zufällig aus der Population gegriffene Person an. Nehmen Sie desweiteren an, dass in der Population einer von 8000 mit Influenza infiziert ist. Stellen Sie Ihren Rechenweg mit allen verwendeten Formeln nachvollziehbar dar und formulieren Sie einen Antwortsatz.

**Aufgabe 4** Für ein Gen, für das vier Allele A, B, C und D unterscheidbar sind, wurden in einer zufälligen Stichprobe von 100 Individuen die verschiedenen Genotypen in folgenden Anzahlen gefunden:

AA	AB	AC	AD	BB	BC	BD	CC	CD	DD
26	11	12	9	10	13	7	4	5	3

- Berechnen Sie das 95%-Wald-Konfidenzintervall für die Häufigkeit des Allels A in der Population.
- Es stellt sich die Frage, ob sich die Population im Hardy-Weinberg-Gleichgewicht befindet. Führen Sie hierzu einen geeigneten  $\chi^2$ -Test durch (inklusive Nullhypothese und Antwortsatz). Sie dürfen dabei verwenden, dass der Wert der Teststatistik gleich  $X^2 = 14.32$  ist. Erläutern Sie aber, was der Genotyp AA zu diesem  $X^2$ -Wert beiträgt.

**Aufgabe 5** Wir kommen noch einmal auf die Sibirischen Tiger zurück.

- Berechnen Sie die Regressionsgerade für die Daten aus Aufgabe 1, um aus dem Gewicht eines Tigers von 2003 vorherzusagen, wie schwer seine heute lebenden Söhne sind.
- Welche Vorhersage macht das Modell für das Gewicht des heute lebenden Sohns eines Tigers, der 2003 ausgewachsen war und 220 kg wog?
- Es stellt sich die Frage, ob beim Gewicht bei Tigern die Vererbung überhaupt eine Rolle spielt oder ob die in den Daten beobachtete Abhängigkeit zwischen den Gewichten der Väter und Söhne auch rein zufällig sein könnten. Führen Sie mit den Daten aus Aufgabe 5 einen geeigneten Test mit Signifikanzniveau  $\alpha = 0.05$  durch. Stellen Sie den Rechenweg nachvollziehbar dar und formulieren Sie Ihr Ergebnis in einem Satz.

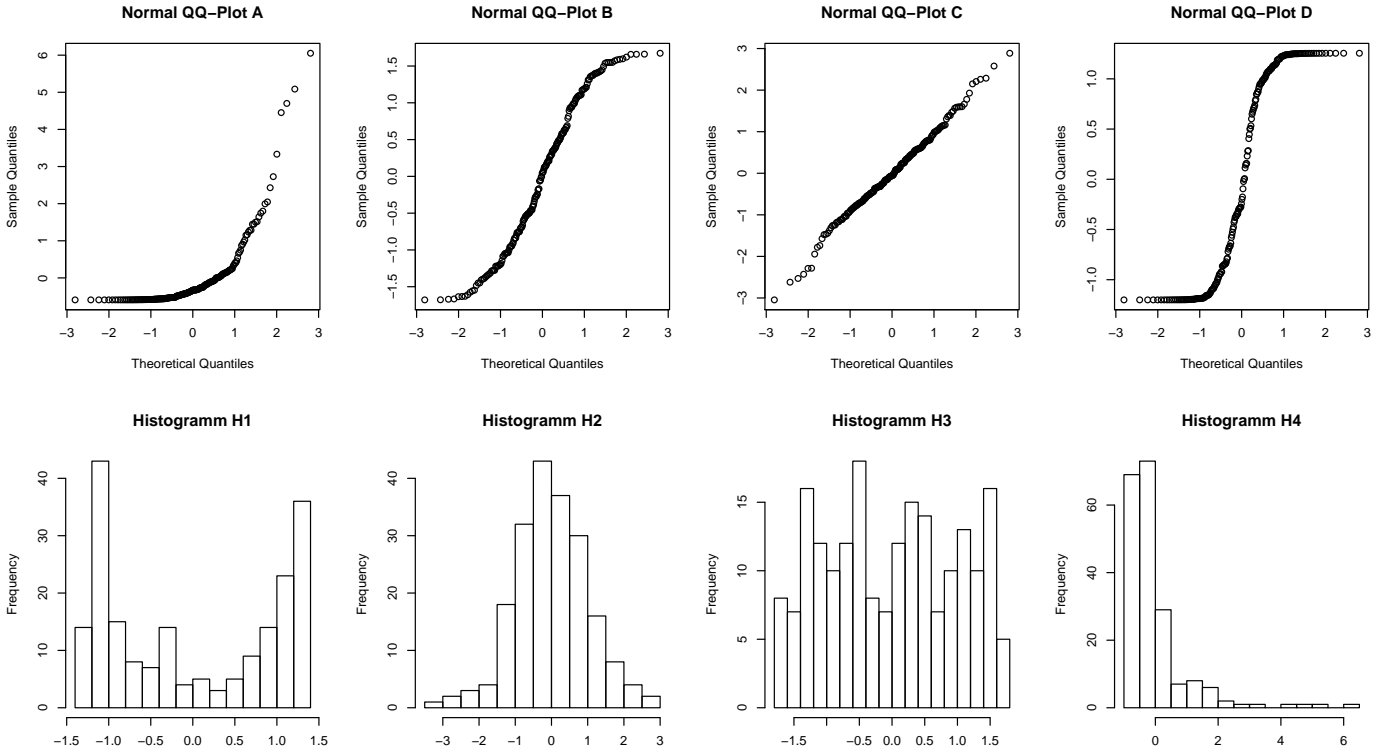
**Aufgabe 6** Pygmäen aus Zentralafrika sind im Durchschnitt 10 cm kleiner als nicht-Pygmäen-Populationen der Region. Sie wollen diesen Unterschied mit einem Test mit Signifikanzniveau 5% und Testmacht 90% zeigen. Wieviele Personen aus jeder Population sollten Sie messen? Aus einem Vorversuch haben Sie die Standardabweichung in beiden Populationen geschätzt: 5.7 cm bei Pygmäen und 6.5 cm bei nicht-Pygmäen.

**Aufgabe 7 – eine online-Aufgabe mit R** In R gibt es einen Beispieldatensatz `PlantGrowth`, der den Ertrag (`weight`) von mehreren Pflanzen bei zwei verschiedenen Behandlungen und einer Kontrollgruppe zeigt. Hier die mit `str(PlantGrowth)` erhaltene Übersicht:

```
'data.frame': 30 obs. of 2 variables:
 $ weight: num 4.17 5.58 5.18 6.11 4.5 4.61 5.17 4.53 5.33 5.14 ...
 $ group : Factor w/ 3 levels "ctrl","trt1",...: 1 1 1 1 1 1 1 1 1 1 ...
```

- Berechnen Sie die Standardabweichung der Erträge von Pflanzen aus der Behandlungsgruppe `trt2`. (Angabe mit drei Nachkommastellen.)
- Welchen p-Wert ergibt ein Welch-t-Test zum Vergleich der mittleren Erträge der Behandlungsgruppen `trt1` und `trt2`? (Angabe mit vier Nachkommastellen.)
- Berechnen Sie den Standardfehler des mittleren Ertrags von Pflanzen aus der Kontrollgruppe. (Angabe mit vier Nachkommastellen.)

**Aufgabe 8** Hier sehen Sie für vier Datenreihen einerseits die Quantile der Daten aufgetragen gegen die Quantile der Normalverteilung und andererseits Histogramme der Daten:



Welcher Quantil-Quantil-Plot gehört zu welchem Histogramm? Vervollständigen Sie die Tabelle:

Normal QQ-Plot	A	B	C	D
Histogramm				